What is new in the redesigned BDS?

The redesigned Business Dynamic Statistics (BDS) data released in Fall 2020 through reference year 2018 reflects improvements and enhancements on several dimensions. This includes a substantial expansion of the set of characteristics over which statistics are released. The most notable are statistics using a consistent NAICS industry classification for the entire time series from 1978-2018 at the sector, 3-digit and 4-digit level and geographic classification at the national, state, MSA and county level. New multi-way interactions include BDS statistics on key variables such as job creation, job destruction, entry and exit at the county by firm size and firm age and MSA by NAICS sector by firm size and firm age groups. This permits tracking the contribution of firm startups to net and gross job creation as well as quantifying the post-entry growth dynamics of young businesses at granular levels not possible in prior releases. As discussed in detail below, the more disaggregate BDS statistics available are possible in part because of the use of noise infusion as the disclosure avoidance methodology. A complete list of tabulations available in the redesigned BDS is in Appendix A.

The redesigned BDS also reflects improvements in source data (especially in early years), an integration with the Statistics of U.S. Businesses (SUSB) data program, greater alignment with County Business Patterns (CBP) data, improvements to the linking methodology and standardization of the production processing. Details of these improvements including those that permit more granular statistics are described as follows:

1. Integration of the best of the SUSB and BDS processing, creating a single set of tables with the greater detail provided by the legacy SUSB employment change tables and the longer time series and firm age dimension found in the BDS.
    - Improved matching algorithms.
    - Improved scope and data quality by combining microdata from both the Business Register (BR) and the County Business Patterns (CBP) throughout the entire time series. Previous releases of the BDS only included CBP-edited data for the last several years of the time series.
2. Linking and scope improvements impacting entry and exit related statistics.
    - An establishment formerly classified as an entrant may now link to the prior year either due to the better matching algorithms or because the matching record was added to the prior year.
    - Similarly, an establishment classified as an exit may now link to the subsequent year.
3. Significant improvements to the algorithms used to retime the bunching of measured establishment entrants and exits within multi-unit firms in economic census years. The new algorithms flexibly incorporate information on inter-censal establishment births and deaths from the Company Organization Survey (COS) into a formal statistical model used to impute first or last year of operation for establishments that appear to be entrants or exits in Census years at multi-unit firms that were not surveyed by the COS. These imputations improve the allocation of the timing of establishment expansions, contractions, entry and exit for multi-unit firms in inter-censal years.

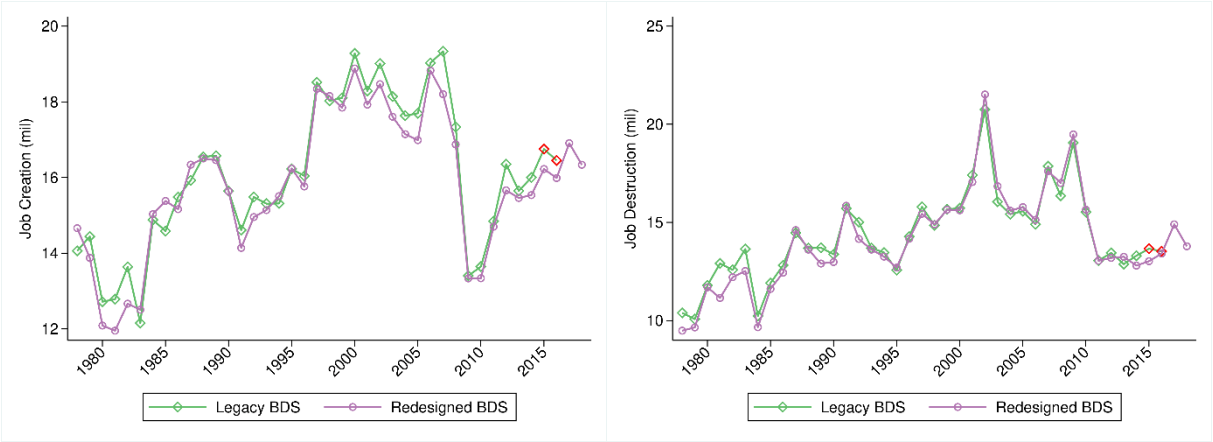
4. All of the improvements described above have been applied to the entire BDS time series.[1] This means that the data quality is more consistent over time than in previous releases of the BDS and statistics are more comparable across years. In particular, entry and exit, job creation from births and job destruction from deaths are measured in the same way over the entire time series, because we now use the same improved linking methodology throughout.
5. Vintage consistent NAICS codes (single classification vintage, e.g. 2012 NAICS codes) have been assigned to all establishments for the entire time series. Using these vintage consistent NAICS codes we are able to produce business dynamics measures for the entire span of the data at significantly more detailed industry breakouts. The vintage consistent algorithms, developed by Fort & Klimek (2018), apply official and derived concordances between different vintages of industry coding schemes (including Standard Industrial Classification codes that were phased out beginning in 1997) to assign a more recent vintage of NAICS codes to establishments observed as far back as 1976. The vintage consistent codes also leverage both partial and longitudinal information to make the most accurate assignment possible when administrative information is limited.
6. In combining the BDS and SUSB data products, we are able to provide measures within detailed geographies found in the SUSB over the longer time series found in the BDS.
7. The new production system also affords increased transparency in how the underlying microdata are created and used to generate the public use BDS data products. Detailed specifications and code used to create the LBD will be available to researchers with approved projects via the FSRDC research network. Additional documentation will be forthcoming as a CES working paper that will describe in detail the production system that creates the LBD microdata and BDS tabulations.
8. Following the SUSB and CBP data products, the new BDS tables use multiplicative noise to avoid the disclosure of sensitive information. See https://www.census.gov/programs-surveys/cbp/technical-documentation/methodology.html for details.

[1] In previous releases of the BDS, improvements to the linking methodology, for example, were only applied to the most recent years of the time series.

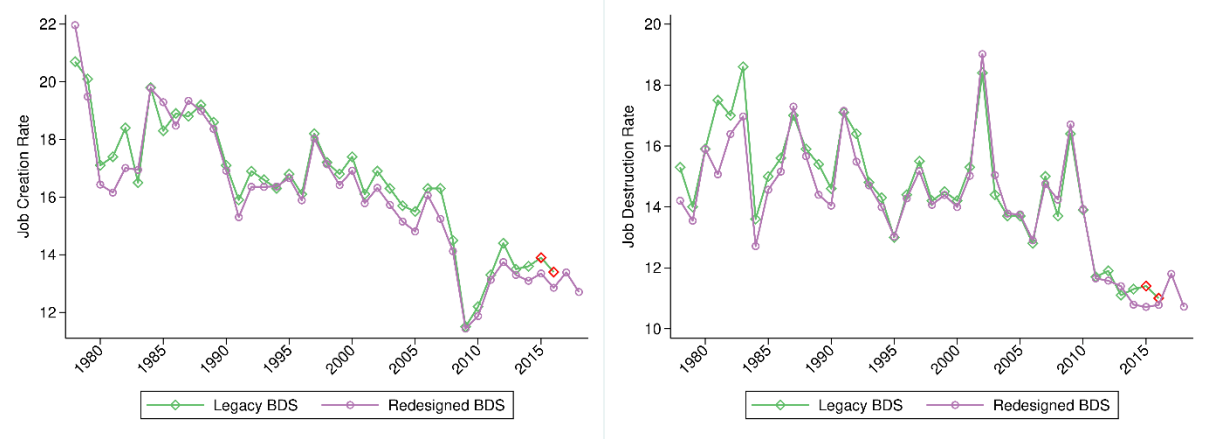
How does the new BDS data compare to the legacy BDS data?

9. Levels and trends in firm, establishment, and gross and net job flows in the new BDS tables are very similar to the legacy tables. Patterns in net job creation line up well across a number of different Census Bureau data products including the legacy BDS tables, new BDS tables, CBP, and SUSB employment change tables, as well as the Bureau of Labor Statistics Business Employment Dynamics data.

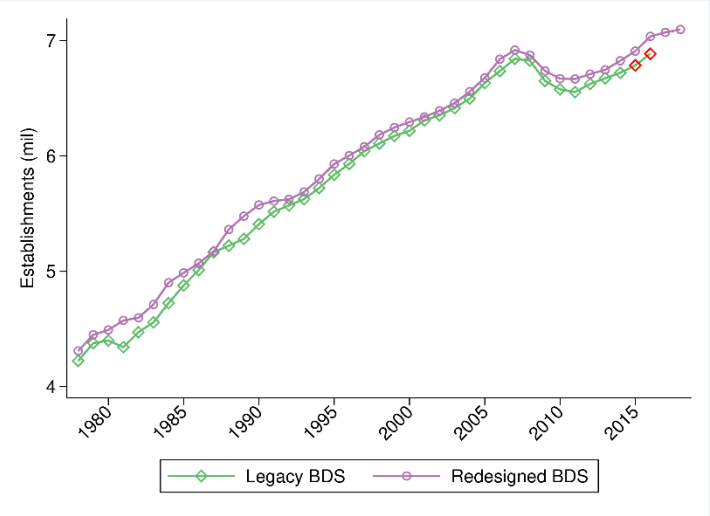
10. The time series pattern of job creation and destruction levels, which capture the job gains from expanding and opening establishments and the job losses from contracting and closing establishments, overlap almost exactly between the new and legacy tables. The legacy series used here is the 2014 release appended with the 2015 and 2016 single-year tabulations, which are marked in red.



Moreover, the job creation and job destruction *rates* are closely aligned in the new and legacy series. Job creation and destruction rate in both series exhibit the familiar secular decline in flows that has been highlighted as an indicator of declining business dynamism.



11. By combining both the BR and CBP microdata, we have increased the number of establishments contributing to the BDS by anywhere from 50,000 to 200,000, depending on the year.

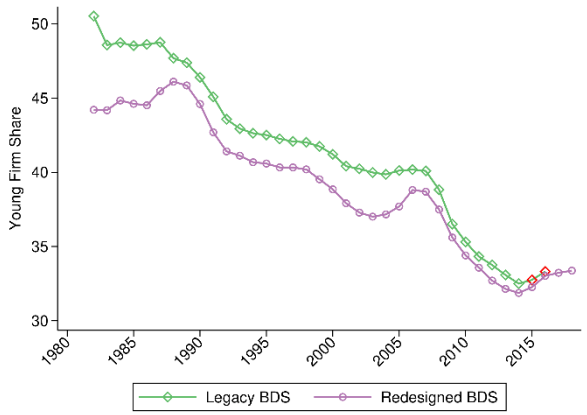


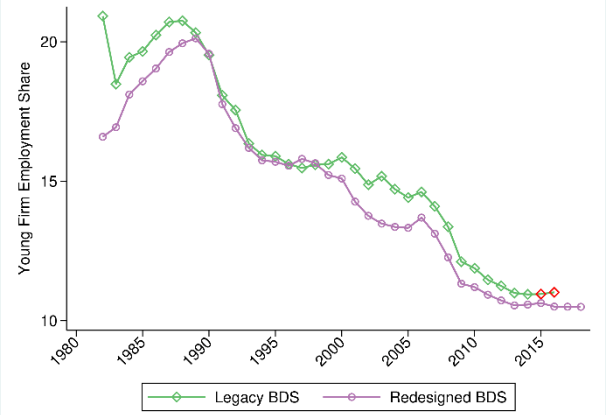
12. Establishment entry and exit rates (the ratio of establishment entrants and exits to all establishments) are somewhat reduced in some years, which reflects both improved linkages and additional data. The entry rate series is notably smoother in the 1980s. Both measures are also close on an employment weighted basis. The job creation rate from births and job destruction rate from deaths align closely in the legacy and new tabulations. The familiar pronounced decline in the entry rate and job creation from births is present in the re-designed data. The establishment exit rate appear more flat through the late 2000s in the new tabulations and then declining through the early 2010s. Job destruction from establishment deaths exhibits greater similarity between the legacy and re-designed series especially after the 1980s. Both the legacy and re-designed series exhibit pronounced declines in the post 2000s period. The improvement in the retiming algorithm as well as improved data is evident in less pronounced spikes in economic census year (i.e., years ending in 2 and 7) entry rates and to a lesser extent in job creation rates from births. The improvement is most notable in 1982, 1997 and 2012. As discussed further below, 2002 remains a challenging year given the re-design of the Census Business register in that year. There is less of an impact from the improvement in the retiming

algorithm in Census year spikes in exit rates (as both a percent of establishments and on an employment-weighted basis).

13. The firm entry rate exhibits a pronounced secular decline in both the re-designed and legacy series. This also holds on an employment weighted basis. There are slightly lower rates in the early 1980s and during the 1990s in the re-designed series. The firm exit rate whether as a share of firms or on an employment-weighted basis are also similar in the redesigned and legacy series. There is much less of a secular decline in exit rates in both series implying that in both there is a pronounced decline in net entry rates of firms. There are some spikes in the re-designed employment-weighted firm exit rates not present in the legacy series, one of the most pronounced being in 2002. As noted above, 2002 is the year of the redesign of the Census Business Register and is a challenging year for longitudinal linking of firms and establishments. In addition, the redesigned system is more efficient at identifying multi-unit firm deaths. As was true in the legacy system, establishment matching using fuzzy name and address algorithms is relatively limited among multi-units. Future research on multi-unit linkage methods may produce linkages that would attenuate the surge of firm death employment in 2002 and perhaps other years as well.

14. The lower firm entry rate in the early 1980s and then later in the 1990s translates into a lower share of both firms and employment coming from young firms (firm age <=5). Even with these differences, the share of young firms in both the legacy and re-designed series declines from about 45% around 1990 to about 32% in 2014. The share of employment in young firms, like other employment-weighted series, is more similar in the legacy and re-designed series. There are lower young firm employment shares in the early 1980s in the re-designed series. The shares of employment at young firms in the legacy and re-designed series are closely aligned in the 1990s. The share at young firms declines more rapidly in the post 2000 period in the redesigned series. In both the legacy and redesigned series there is a sharp decline in Great Recession which persists in the newly released series through 2018.


Young Firm Share
50
45
40
35
30
1980
1985
1990
1995
2000
2005
2010
2015
Legacy BDS
Redesigned BDS


Young Firm Employment Share
20
15
10
1980
1985
1990
1995
2000
2005
2010
2015
Legacy BDS
Redesigned BDS

Other detailed differences

15. Fewer detailed size groupings: The more aggregated employment size groupings from the SUSB are used throughout the new BDS tables. The groups "20 to 49" and "50 to 99" were combined and "100 to 249" and "250 to 499" were combined.
16. Following the SUSB, tables with greater geographic detail that provide measures by size or age use more aggregate size and age groupings.
17. The new tables are no longer organized by "firm" and "establishment" characteristics tables. Some characteristics are measured at the firm-level and some are measured at the establishment-level. Establishment-level characteristics include industry, geography, and establishment size and establishment age. Firm level characteristics include firm size, initial firm size, and firm age. The tables sometimes include one or both types of characteristics.
18. The new BDS begins measuring firm, establishment, and employment stocks and flows in 1978. We chose this starting year because 1977 is the first economic census year in the LBD microdata and hence serves as a good baseline of existing establishments. Also since there is only one pre-1977 year (1976), the retiming of economic census year births and deaths cannot be done in 1977 as it is done for subsequent economic census years. As such, the first high-quality year-to-year changes we observe are from 1977 to 1978.
19. The initial release using the vintage consistent industry codes will not include tabulations for Agriculture, Forestry, Fishing, and Hunting (NAICS 11). Industries 111 Crop Production and 112 Animal Production and Aquaculture are out of scope for both the BDS and CBP and hence are not included in any BDS tables. Industries 113, 114, and 115 (Forestry, Fishing and Hunting, and Support Activities for Agriculture and Forestry respectively) are in scope for both the BDS and CBP and are included in all tables that are not broken out by industry. However, NAICS 11 and the corresponding in-scope 3 and 4 digit industries are excluded from the industry breakout tables. This is partly due to limitations of the current VCNAICS algorithm, which does not assign detailed NAICS codes to establishments in NAICS 11, and partly due to problems with consistency across the long BDS time series in the agriculture sector. The Census Bureau is actively working to resolve these issues and plans to publish 2, 3, and 4 digit industry tables for the agriculture sector in the near future.
20. As was done in the legacy BDS tables, time invariant geography codes are used in the establishment-level microdata. That is, establishments are assigned unique geography codes (state, county, MSA, etc.) for all years that they contribute to the tabulations. This is done by using the most recent Economic Census year, or the most frequently assigned code if an establishment was not observed in any Economic Census. We preference Economic Census years because the Economic Censuses tends to provide the most accurate establishment level information. We also use contemporaneously assigned geography codes without applying longitudinal consistency checks based on changes in code definitions. This was not a concern in the legacy BDS tabulations since state was the most disaggregate geography code and state definitions are stable over time. County codes and definitions, in contrast, do change over time. The interaction of these two concepts, time invariant assignment of geographic codes and changing county definitions, should be kept in mind when comparing county-level measures over long periods of time.
21. In spite of extensive quality assurance work on the new data, there remain some odd patterns in the BDS tables that cause concern about measurement error. When we deem any statistic in the

highly aggregated tables—i.e. only tables with 1 by-variable or fewer—to be too uncertain because of very large changes across years (i.e. spikes), we suppress this statistic and report only '(S)' in the cell. More specifically, these large spikes represent cases where a statistic sees a large change in one direction (positive or negative) in a given year, followed by a similarly large change in the opposite direction in the following year. Large year to year changes are identified in terms of their absolute value and percentage change, as well as their relative significance in the context of the time series.

22. The new BDS statistics are based on noisy employment, meaning we add a noise factor to the employment for every establishment before doing statistical calculations. This new disclosure protection system is another reason why legacy BDS numbers will not be identical to new BDS numbers even in years that were previously published.
23. The convention for computing DHS denominator for the count of establishments in computing the entry rate changed in the redesigned BDS, which results in an establishment entry rate for age 0 establishments and firms equal to 100 rather than 200 in the legacy BDS. In the re-designed BDS, an establishment counts as one in both t-1 and t computation of the DHS denominator as long as the establishment has positive Denom=$0.5*(Emp_{t-1}+Emp_t)$. The DHS denominator for establishment entry rate for any cell is equal to $0.5*(Estabs_{t-1}+Estabs_t)$ where $Estabs_t$ is the count of establishments in the cell in year t. Since the new convention is to count entrants as contributing to the establishment count for measuring this DHS denominator for entry rate in both t-1 and t, this implies that the typical entry rate for age 0 is 100. For example, if age 0 has 50 entrants then the numerator of the entry rate is 50 and the denominator is 0.5*(50+50) so the entry rate (multiplied by 100) is 100. In the legacy BDS an entering establishment did not count in the DHS denominator for the entry rate in year t-1. For the above example, the numerator would be 50 and the denominator 25 so the entry rate (multiplied by 100) is 200. Note that this change in convention only has noticeable implications for age 0 cells for establishment entry rates. Similar remarks apply to firm entry rates in the firm age = 0 cell. For the employment-weighted entry rate (i.e., job creation rate from births) the convention is the same in both the legacy and redesigned BDS. The DHS denominator used for the job creation rate from births is Denom and $Emp_{t-1}=0$ for entrants. Hence the job creation rate for births in the establishment age 0 cell is 200 in both the legacy and re-designed BDS.

Appendix A

List of classifications with BDS statistics released:

Economy Wide

year

One-way Tables

year, bds_fage
year, bds_eage
year, bds_fsize
year, bds_esize
year, bds_ifsize
year, bds_iesize
year, bds_metro
year, bds_st
year, bds_msa
year, bds_cty
year, bds_sector
year, bds_vcnaics3
year, bds_vcnaics4

Two-way Tables

year, bds_fage, bds_fsize
year, bds_fage, bds_ifsize
year, bds_eage, bds_esize
year, bds_eage, bds_iesize
year, bds_st, bds_sector
year, bds_st, bds_metro
year, bds_st, bds_fage
year, bds_st, bds_fsize
year, bds_st, bds_ifsize
year, bds_st, bds_eage
year, bds_st, bds_esize
year, bds_st, bds_iesize
year, bds_sector, bds_fage
year, bds_sector, bds_fsize
year, bds_sector, bds_ifsize
year, bds_sector, bds_eage
year, bds_sector, bds_esize
year, bds_sector, bds_iesize
year, bds_vcnaics3, bds_fage
year, bds_vcnaics3, bds_fsize

year, bds_vcnaics3, bds_ifsize
year, bds_vcnaics3, bds_eage
year, bds_vcnaics4, bds_fage
year, bds_vcnaics4, bds_fsize
year, bds_vcnaics4, bds_ifsize
year, bds_vcnaics4, bds_eage
year, bds_metro, bds_fage
year, bds_metro, bds_fsize
year, bds_metro, bds_ifsize
year, bds_metro, bds_eage
year, bds_metro, bds_esize
year, bds_metro, bds_iesize
year, bds_metro, bds_sector
year, bds_msa, bds_fagecoarse
year, bds_msa, bds_fsizecoarse
year, bds_msa, bds_ifsizecoarse
year, bds_msa, bds_eagecoarse
year, bds_msa, bds_sector
year, bds_cty, bds_fagecoarse
year, bds_cty, bds_fsizecoarse
year, bds_cty, bds_ifsizecoarse
year, bds_cty, bds_eagecoarse
year, bds_cty, bds_sector

Three-way Tables

year, bds_msa, bds_sector, bds_fagecoarse
year, bds_msa, bds_sector, bds_fsizecoarse
year, bds_msa, bds_sector, bds_ifsizecoarse
year, bds_msa, bds_sector, bds_eagecoarse
year, bds_st, bds_sector, bds_fage
year, bds_st, bds_sector, bds_fsize
year, bds_st, bds_sector, bds_ifsize
year, bds_st, bds_sector, bds_eage
year, bds_metro, bds_sector, bds_fage
year, bds_metro, bds_sector, bds_fsize
year, bds_metro, bds_sector, bds_ifsize
year, bds_metro, bds_sector, bds_eage
year, bds_metro, bds_st, bds_fage
year, bds_metro, bds_st, bds_fsize
year, bds_metro, bds_st, bds_ifsize
year, bds_metro, bds_st, bds_eage
year, bds_metro, bds_st, bds_sector

Four-way Tables

year, bds_st, bds_metro, bds_sector, bds_fage

year, bds_st, bds_metro, bds_sector, bds_fsize

year, bds_st, bds_metro, bds_sector, bds_ifsize

year, bds_st, bds_metro, bds_sector, bds_eage